

Embracing events in causal modelling: Interventions and counterfactuals in CP-logic

Joost Vennekens^{*1}, Maurice Bruynooghe², and Marc Denecker²
firstname . lastname @ cs.kuleuven.be

¹ Campus De Nayer, Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium

² Dept. Comp. Sc., K.U. Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium

Abstract. This paper integrates Pearl’s seminal work on probability and causality with that of Shafer. Using the language of CP-logic, it transposes Pearl’s analysis of interventions and counterfactuals to the semantic context of Shafer’s probability trees. This gives us definitions that work not on the level of random variables, but on the level of Humean events. There are some tangible benefits to our approach: we can elegantly handle counterfactuals in the context of cyclic causal relations, and are able to consider interventions that are both more fine-grained and more elaborate than Pearl’s.

1 Introduction

Causal statements implicitly refer to things that happen. For instance, the statement “syphilis causes paresis” refers to the biological process of syphilis spirochaetes damaging certain brain cells. It is by means of this process that patients who at first exhibit only a syphilis infection will eventually come to exhibit paresis as well. The causal statement itself leaves the details of this process implicit and just asserts its existence: “somehow,” it says, “syphilis causes paresis.” Each causal statement of this form (e.g., dropping a vase causes it to break, a voltage drop causes an electrical current, being born in Belgium causes Belgian citizenship) implicitly refers to some such implicit process, to some implicit thing that happens to generate the effect from its cause.

In [9], Shafer recognizes the importance of this dynamic aspect of causality and introduces a specific term for these implicit things-that-happen: *Humean events*. The adjective “Humean” is added to avoid confusion with the technical meaning that the term “event” has in probability theory. That is, a Humean event is *not* a subset of some sigma algebra, nor a set of possible outcomes of some experiment, but simply, as in everyday language, something that happens. Following the line of thought from the preceding paragraph, Shafer places this concept at the centre of his work, using probability trees to model causal systems as sequences of Humean events. In this context, a random variable (RV), for instance, is simply something which gets assigned a certain value at some point in the sequence of events.

^{*} Partially supported by IWT-Vlaanderen.

A different approach is taken by Pearl [7]. His probabilistic causal models (PCMs) represent the relations between RVs as they hold in stable states of the domain (i.e., when no more Humean events are happening). His models, therefore, do not care about which event causes the value of an RV, or when this happens. For instance, [5] uses the term “event” to refer to an assignment $\mathbf{X} = \mathbf{x}$ of values to RVs. Unlike proper Humean events, such assignments are not things that intrinsically happen at a certain point in time, as part of a larger sequence of events. Underlying one such an assignment, there may be many different Humean events, happening at different times, and it may even depend on the context (i.e., the outcome of previous events) which events are involved in a particular assignment. For instance, a patient being in hospital ($Hospital = true$) may be the result of a car crash yesterday (a Humean event), but also, if he managed to avoid the crash, of slipping on the stairs this morning (a different Humean event). Shafer’s account therefore presents a more refined view on the dynamic aspects of causality than Pearl’s. However, it lacks Pearl’s thorough treatment of *interventions* and *counterfactuals*.

In [10] (and before at JELIA 2006), we presented a logical language called *CP-logic* (the “CP” stands for causal and probabilistic), which offers a succinct syntax for describing classes of Shaferian probability trees. We will now use this language to provide the semantics of probability trees with a suitable notion of interventions. In the same way as Shafer’s account of causality in terms of Humean events can be seen as a refinement of the RV-based model employed by Pearl, our account of interventions and counterfactuals in CP-logic will constitute a useful refinement of the account found in [7].

2 Preliminaries: CP-logic

We assume familiarity with classical logic and briefly recall CP-logic. For simplicity, we omit function symbols and make the Herbrand assumption of identifying interpretations with sets of ground atoms. More detail can be found in [10].

A *causal probabilistic law*, or *CP-law* for short, is a statement r of the form:

$$\forall \mathbf{x} (A_1 : \alpha_1) \vee \cdots \vee (A_n : \alpha_n) \leftarrow \phi.$$

Here, ϕ is a first-order formula, the A_i are atoms, the α_i are non-zero probabilities, and the tuple of variables \mathbf{x} contains all free variables in ϕ and the A_i . Intuitively, for each \mathbf{x} , $\phi(\mathbf{x})$ causes some implicit Humean event, which will result in at most one of the effects $A_i(\mathbf{x})$. For each i , α_i is the probability of A_i being the resulting effect. Therefore, we require $\sum \alpha_i \leq 1$. If the equality holds, exactly one A_i is caused; otherwise, it is possible that the Humean event passes without any (visible) effect on the state of the world. For mathematical uniformity, we introduce the notation $r^=$ to refer to r itself if the equality holds, and otherwise to the CP-law: $\forall \mathbf{x} (A_1 : \alpha_1) \vee \cdots \vee (A_n : \alpha_n) \vee (\text{—} : 1 - \sum_i \alpha_i) \leftarrow \phi$. Here, the dash represents the possibility of there being no (visible) effect. For a CP-law r , we denote \mathbf{x} by $vars(r)$, refer to ϕ as $body(r)$, to the sequence $(A_i, \alpha_i)_{i=1}^n$ as $head(r)$, and to $(A_i)_{i=1}^n$ as $head_{At}(r)$. We also allow the condition $body(r)$ to be

omitted, which means that the Humean event always happens. We abbreviate $(A : 1) \leftarrow \phi$ by $A \leftarrow \phi$. A *CP-theory* is a finite set of CP-laws.

We now recall the semantics of such a CP-theory. For simplicity, we will assume that each precondition $body(r)$ is a positive formula. The semantics of negation can be found in [10].

Our basic semantic construct is a probability tree [9], i.e., a finite tree in which each edge is labeled with a probability, such that the labels of all edges leaving the same internal node always sum up to one. Intuitively, such a tree \mathcal{T} represents a probabilistic process: each node is a state of the process and, together, the edges leaving a node represent a Humean event that causes a probabilistic transition to one of its children. The root is the initial state and the leaves are final states. Let us denote by $\pi_{\mathcal{T}}$ the probability distribution that a tree \mathcal{T} defines over its leaves, i.e., for each leaf l , $\pi_{\mathcal{T}}(l)$ is the product of the labels of the edges that lead to l . We now define which probability trees \mathcal{T} correspond to a given set of CP-laws C . Basically, the events that happen in \mathcal{T} should follow the blueprints given by the CP-laws in C . An *occurrence* of a CP-law $r \in C$ is the result $r[\mathbf{x}/\mathbf{c}]$ of replacing the variables $\mathbf{x} = vars(r)$ by constants \mathbf{c} . The *grounding* $grnd(C)$ of C consist of all occurrences of CP-laws $r \in C$ that can be thus constructed. Because we have no function symbols and only a finite set of constants, this grounding is finite.

Condition 1 *Events correspond to occurrences of CP-laws, i.e., there exists a mapping \mathcal{E} from the internal nodes s of \mathcal{T} to $grnd(C)$ such that we can label the nodes and the edges of \mathcal{T} as: if $\mathcal{E}(s) = r[\mathbf{x}/\mathbf{c}]$ and $head((r[\mathbf{x}/\mathbf{c}])^\circ) = (A_i, \alpha_i)_{i=0}^n$, then the children of s are nodes s_0, \dots, s_n such that the label $lbl(s_i) = A_i$ and $lbl((s, s_i)) = \alpha_i$. For uniformity, the root is labeled with ‘—’ (“no-effect”).*

Condition 2 *An occurrence that has already happened cannot happen again. Formally, let $Anc(s)$ be the ancestors of s (not including s itself), and let $\mathcal{R}(s)$ be the set of occurrences that have not yet happened in s (i.e., $\mathcal{R}(s) = grnd(C) \setminus \{\mathcal{E}(s') \mid s' \in Anc(s)\}$). Then each $\mathcal{E}(s)$ must be in $\mathcal{R}(s)$.*

Finally, we relate the nodes in the tree to the effects and preconditions of the occurrences. Recall that each node s in a probability tree corresponds to a potential state of the domain. We represent this state by an interpretation (i.e., a set of ground atoms) as follows: $\mathcal{I}(s) = \bigcup_{t \in Anc(s) \cup \{s\}} l(t)$, where $l(t) = \{\}$ if $lbl(t) = \text{‘—’}$ and $\{lbl(t)\}$ otherwise.

Condition 3 *The precondition of the occurrence of a CP-law that happens in a node must be satisfied in its state, i.e., for all internal nodes s : $\mathcal{I}(s) \models body(\mathcal{E}(s))$.*

The three conditions above describe when a probability tree unfolds according to the CP-laws of a theory C . We call such a tree an *execution model* of C if it is also complete in the following sense:

Condition 4 *\mathcal{T} cannot be extended, i.e., for each leaf l and $r \in \mathcal{R}(l)$, we have that $\mathcal{I}(l) \not\models body(r)$.*

While we lack space for an example at this point, Fig. 1 on page 6 show an execution model for Example 1. By defining a probability distribution $\pi_{\mathcal{T}}$ over its leaves, an execution model \mathcal{T} induces a probabilistic possible world semantics: the probability $\pi_{\mathcal{T}}(S)$ of an interpretation S is $\sum_{\mathcal{I}(l)=S} \pi_{\mathcal{T}}(l)$. The probability of a formula ϕ is then $\pi_{\mathcal{T}}(\phi) = \sum_{S \models \phi} \pi_{\mathcal{T}}(S)$. In [10], it was shown that each execution model \mathcal{T} of a CP-theory C defines the *same* possible world semantics $\pi_{\mathcal{T}}$, which we therefore also denote as π_C . This result demonstrates one sense in which causal statements are indeed justified in leaving implicit the Humean events to which they refer: as long as we are only interested in the final states that will eventually be reached (and not in any intermediate states), the properties of these implicit events do not matter.

CP-logic normally distinguishes exogenous from endogenous predicates [10]. Here, however, we will save some space by writing “ $\forall \mathbf{x} (Exo(\mathbf{x}) : *) \leftarrow$ ” to say that *Exo* should really be exogenous, but that we will treat it as an endogenous predicate caused with some unknown probability $*$. For the examples in this paper, it is easy to see that the proper treatment of exogenous predicates would yield the same results.

CP-logic is closely related to Probabilistic Causal Models (PCMs). A PCM that contains only boolean equations with boolean RVs can easily be translated to a CP-theory that contains for each PCM equation a single CP-law that propagates the value of the body of the equation to its head. Conversely, CP-theories can also be translated to PCMs. If the CP-theory contains no cycles (to be discussed below), then this is trivial. Otherwise, artificial RVs representing common causes underlying all of the variables in a cycle are needed. See [10] for details.

3 Interventions

The goal of this section is to provide CP-logic with definitions that capture Pearl’s intuitions about interventions. Let us briefly recall some of the formal tools used by Pearl. A *structural model* is a set of *structural equations*, each of which defines the value of one RV in terms of the values of some other RVs. A *Probabilistic Causal Model (PCM)* consists of a structural model, together with a probability distribution over the values of its exogenous RVs (i.e., those without a defining equation). An important restriction is that, given any assignment of values to the exogenous RVs, this set of equations must have a unique solution. Typically, this is ensured by requiring an acyclic set of equations.

An *intervention* in a PCM is of the form $do(\mathbf{X} = \mathbf{x})$ with \mathbf{X} of a tuple of endogenous variables and \mathbf{x} a tuple of values. *Performing* this intervention means replacing the defining equation of each $X_i \in \mathbf{X}$ by $X_i := x_i$. In a logical framework, we can view a PCM as an acyclic set of equations $A := \phi$ where A is a ground atom and ϕ a sentence. An intervention $do(\mathbf{A} = \mathbf{a})$ then assigns a tuple of truth values $\mathbf{a} \in \{\mathbf{t}, \mathbf{f}\}^n$ to the tuple of atoms \mathbf{A} , while removing their defining equations. In this paper, we will not consider interventions at the level of random variables (i.e., atoms) as Pearl does, but look instead at interventions that add and/or prevent CP-laws to/in a theory. Because CP-logic represents a

causal system in a modular way as a set of causal laws, this kind of intervention is already built into its semantics.

Definition 1. Let C be a CP-theory. An intervention is a pair (R, A) with R a subset of C (a preemption) and A a set of laws not in C (an addition). The result of performing (R, A) on C , denoted $C \downarrow (R, A)$, is the CP-theory $(C \setminus R) \cup A$.

Let r be $\forall \mathbf{x} (A_1 : \alpha_1) \vee \dots \vee (A_n : \alpha_n) \leftarrow \phi$. For atoms $\mathbf{A} \subseteq \{A_i \mid 1 \leq i \leq n\}$, let $r|_{\bar{\mathbf{A}}}$ be this CP-law without the atoms \mathbf{A} , i.e.,

$$r|_{\bar{\mathbf{A}}} = \forall \mathbf{x} \bigvee_{A_i \notin \mathbf{A}} (A_i : \alpha_i) \leftarrow \phi.$$

Let $r|_{A_i}^+$ be the CP-law $A_i \leftarrow \phi$. An intervention that blocks only the possible effect A_i can be represented as $(\{r\}, \{r|_{\bar{\mathbf{A}}}\})$. At the other extreme, $(\{r\}, \{r|_{A_i}^+\})$ forces the outcome of the event caused by ϕ to be A_i .

We can use this notion of intervention to simulate Pearl's interventions. A Pearl-style intervention $Int = do(A_0 = \mathbf{t}, \dots, A_m = \mathbf{t}, B_0 = \mathbf{f}, \dots, B_n = \mathbf{f})$ corresponds to a pair (R, A) where $R = \bigcup_{j \geq 0} \{r \in C \mid B_j \in head_{At}(r)\}$ and $A = \bigcup_{j \geq 0} \{r|_{\{B_j \mid 0 \leq j \leq n\}} \mid r \in R\} \cup \bigcup_{i \geq 0} \{A_i \leftarrow \text{"}\}$. Given a PCM P and its corresponding CP-theory C (as in [10]), performing intervention Int on P produces the same probability distribution as performing the intervention (R, A) on P . (Proof omitted because of space restrictions.)

4 Counterfactuals

Let us again start by recalling Pearl's treatment of counterfactuals. He considers the following class of statements:

$$\underbrace{\text{Given } X}_{\text{explanation}}, \underbrace{\text{would } Y \text{ have happened}}_{\text{prediction}}, \underbrace{\text{had we done } Z}_{\text{intervention}}?$$

Pearl's intuition is to read such a statement as: if we do the intervention Z , will Y then hold, assuming that, insofar as the intervention does not interfere, whatever lead to X in the first place will still happen in the same way as it did before? He therefore suggests the following three-step process for evaluating such statements in a PCM.

Explanation: update the *a priori* distribution over the exogenous variables by the observation X .

Intervention: apply the intervention Z to the model.

Prediction: compute the probability of Y in the resulting model, using the *a posteriori* distribution given X , i.e., we look at $P(Y \mid do(Z), X)$.

Again, our goal is to see how we can apply these intuitions in the context of CP-logic. Let us introduce our approach with an example from [7].

Example 1. A court might order the death of a prisoner. The probability of this is p . The execution is to be performed by a two person firing squad. If the court so decides, the captain of the firing squad orders both of his riflemen to fire. However, rifleman A is of the nervous type, and might shoot even if not ordered to. This happens with probability q . If at least one rifleman fires, the prisoner dies. In CP-logic, this becomes:

$$(Court : p) \leftarrow \quad (1)$$

$$Capt \leftarrow Court \quad (2)$$

$$Fires(A) \leftarrow Capt \quad (3)$$

$$(Fires(A) : q) \leftarrow \neg Capt \quad (4)$$

$$Fires(B) \leftarrow Capt \quad (5)$$

$$Death \leftarrow Fires(A) \quad (6)$$

$$Death \leftarrow Fires(B) \quad (7)$$

Note that (4) contains negation, the semantics of which was not explained in Section 2 but can be found in [10]. We could also omit $\neg Capt$ from this CP-law; the only effect would be that $Fires(A)$ might redundantly be caused twice (in branch l_2 of Fig. 1a), which does not change the semantics of this theory.

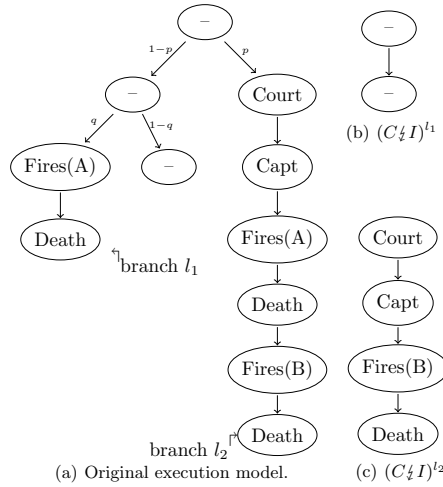


Fig. 1. Execution models of example 1.

the *a posteriori* probability $P(l_1 \mid Death)$ is $q(1-p)/n$ and $P(l_2 \mid Death) = p/n$ with $n = q(1-p) + p$.

Intervention: To prevent A from firing, we apply the intervention that removes CP-laws (3) and (4). Now, as shown in Fig. 1b, if all the remaining events happen in the same way as they happened in l_1 , the prisoner survives; on the other hand, as Fig. 1c shows, if they happen as they happened in l_2 , he still dies.

Prediction: Having thus determined that the intervention would have prevented the prisoner's death just in case the original execution model was actually in branch l_2 instead of l_1 , we can now judge the desired probability to be $\pi_{\mathcal{T}}(l_2 \mid Death)$, which is p/n .

Pearl considers the counterfactual question: "If the prisoner is dead, what is the probability that he would still be dead if rifleman A had not shot?" Intuition puts it at the probability that the prisoner's death is due to the court ordering his execution, i.e., at $P(Court \mid Death) = \frac{p}{q(1-p)+p}$, because it is precisely in this case that the intervention of preventing A from firing will not save the prisoner (since B will also fire). We can reach this conclusion with a CP-logic variant of Pearl's three steps, starting from the execution model \mathcal{T} in Fig. 1a.

Explanation: \mathcal{T} defines a probability distribution $\pi_{\mathcal{T}}$ over its leaves. Having observed $Death$, we can update this distribution: because l_1 and l_2 are the branches where the prisoner dies,

To make this more formal, we need a number of definitions. Let us choose an execution model \mathcal{T} for our theory (it can be shown that the choice does not affect the outcome). In the explanation step, we condition on the observation X as follows.

Definition 2. For a formula X , we define the conditional probability $\pi_{\mathcal{T}}(l \mid X)$ of a leaf l of \mathcal{T} given X to be 0 if $\mathcal{I}(l) \not\models X$ and $\pi_{\mathcal{T}}(l)/\pi_{\mathcal{T}}(X)$ otherwise.

In the intervention step, we consider what would have happened in different circumstances, under the assumption that all events that are not affected would still have happened in the same way as they originally happened.

Definition 3. Let l be a leaf of an execution model of a CP-theory C and $B(l)$ the branch that leads to l . By C^l we denote the CP-theory that fixes the outcome of all the events that happened in $B(l)$ to the outcome they had in $B(l)$:

$$C^l = \{r \in C \mid r \in \mathcal{R}(l)\} \cup \{r|_H^+ \mid \exists s \in B : \text{lbl}(s) = H \text{ and } \mathcal{E}(\text{parent}(s)) = r\}.$$

For instance, if we fix the outcomes that occurred in l_1 :

$$C^{l_1} = \left\{ \begin{array}{lll} - \leftarrow, & \text{Fires}(A) \leftarrow \text{Capt}, & \text{Death} \leftarrow \text{Fires}(A), \\ \text{Capt} \leftarrow \text{Court}, & \text{Fires}(A) \leftarrow \neg \text{Court}, & \text{Death} \leftarrow \text{Fires}(B), \\ & \text{Fires}(B) \leftarrow \text{Capt}, & \end{array} \right\}$$

We now define counterfactual probabilities as follows.

Definition 4. Let X, Y be formulas (respectively observation and prediction) and Z an intervention. The counterfactual probability of Y after Z given X is

$$\text{Cfl}(X, Y, Z) = \sum_{l \text{ is leaf of } \mathcal{T}} \pi_{(C \dot{\downarrow} Z)^l}(Y) \cdot \pi_{\mathcal{T}}(l \mid X).$$

In our example, performing the intervention $Z = (\{(3), (4)\}, \{\})$ on the theory produces the following $C \dot{\downarrow} Z$:

$$\left\{ \begin{array}{lll} (\text{Court} : p) \leftarrow, & & \text{Death} \leftarrow \text{Fires}(A), \\ \text{Capt} \leftarrow \text{Court}, & \text{Fires}(B) \leftarrow \text{Capt}, & \text{Death} \leftarrow \text{Fires}(B) \end{array} \right\}$$

In $(C \dot{\downarrow} Z)^{l_1}$, the first of these CP-laws reduces to “ $- \leftarrow$ ”, whereas in $(C \dot{\downarrow} Z)^{l_2}$, it reduces to “ $\text{Court} \leftarrow$ ”. The reader can verify that the branch in Fig. 1b indeed corresponds to $(C \dot{\downarrow} Z)^{l_1}$ and that Fig. 1c corresponds to $(C \dot{\downarrow} Z)^{l_2}$. So,

$$\text{Cfl}(\text{Death}, \text{Death}, Z) = 0 \cdot \frac{q(1-p)}{q(1-p)+p} + 0 + 1 \cdot \frac{p}{q(1-p)+p}.$$

5 Causal cycles

Let us now amend the previous example, by adding that each of the two soldiers will also fire when he hears the guy next to him fire. In CP-logic, this is a pretty innocuous change; we simply add:

$$\text{Fires}(A) \leftarrow \text{Fires}(B). \quad \text{Fires}(B) \leftarrow \text{Fires}(A).$$

The only effect of this change is that now two soldiers will fire in circumstances where previously only one would, which raises the probability of $Fires(B)$ from p to $p + (1 - p)q$. Because one soldier firing already suffices to kill the prisoner, however, this does not affect the probability of his death.

The purpose of this section is to demonstrate that such cyclic causality cannot be adequately handled in Pearl’s framework. To this end, we will examine a number of ways of trying to do this, and discuss what is wrong with each of them.

First, let us note that it obviously does not suffice to leave the equation for $Fires(A)$ untouched and change only the equation for $Fires(B)$ into:

$$Fires(B) := Captain \vee Fires(A).$$

Indeed, when we intervene with $Fires(B)$, for instance by sabotaging his rifle so that it will go off even without the soldier pulling the trigger, the effect should be that A also fires, which the current equation for $Fires(A)$ will not accomplish. One is therefore tempted to also make this change:

$$Fires(A) := Caption \vee Nervous \vee Fires(B).$$

Together with our modification of the equation for $Fires(B)$, however, this clearly violates the acyclicity restriction. In the appendix to [4], Halpern and Pearl present a way of lifting this restriction. However, their semantics is not what is needed for this example: they impose an equilibrium condition, where every assignment that satisfies all equations is considered possible. Therefore, it is possible for A to fire for no reason than that B does, and at the same time for B to fire for no other reason than that A does. Clearly, this is not what we want for this example: if the soldiers fire, then at least one of them should have a reason for firing that is not his comrade firing first.

This problem with the semantics of course has consequence for the results that are produced. For instance, even the tautological counterfactual “given that the prisoner survived, he would have survived” cannot be deemed true. Moreover, it also becomes impossible to judge the probability of the prisoner dying any more accurate than that it must be somewhere in the interval $[p + (1 - p)q, 1]$.

The transformation from CP-logic to Bayesian networks given in [10] would attempt to solve this problem by introducing an intermediate RV $BothFire$, replacing the equations for $Fires(A)$ and $Fires(B)$ by:

$$\begin{array}{ll} BothFire := Nervous \vee Captain & Fires(A) := BothFire \\ & Fires(B) := BothFire \end{array}$$

This works insofar as that it generates the right probability distribution for $Death$, but it breaks down when intervention come into play. The reason is of course that this model has removed the asymmetry between $Fires(A)$ and $Fires(B)$: it no longer has the information that A , and not B , is the soldier who might fire out of nervousness. Consequently, if we ask: “given that the captain did not give the order to fire, would B have fired if we had prevented A from

firing?”, then the above model has no way of knowing that the answer should be “no”.

We are therefore forced into more complicated options, such as including multiple copies of our original RVs. For instance, we can have $Fires_i(A)$ and $Fires_i(B)$ for $i \in \{1, 2\}$:

$$\begin{aligned} Fires_2(A) &:= Fires_1(A) \vee Fires_1(B) & Fires_1(A) &:= Captain \vee Nervous \\ Fires_2(B) &:= Fires_1(B) \vee Fires_1(A) & Fires_1(B) &:= Captain \end{aligned}$$

We can think here of the indices as a timestamp for the RVs, which explicitly encodes the small time delay between hearing your neighbour firing and firing yourself. In general, for a firing squad of n soldiers, in which a soldier firing causes his two neighbours to fire too, we would need n copies of each RV, allowing a “falling domino”-style propagation through the squad. An intervention such as preventing soldier S from firing should then be interpreted as an intervention with all RVs S_i , $i = 1..n$.

This solution is correct, but has the downside of blowing up the representation: simply adding the domino-effect forces us to abandon the original representation in terms of n RVs, in favour of a new representation in term of n^2 RVs. This is neither concise, nor elaboration tolerant. By contrast, in CP-logic, one just needs to add the obvious n CP-laws, in terms of the original vocabulary:

$$\begin{aligned} Fires(S_1) &\leftarrow Fires(S_2). \\ &\dots \\ Fires(S_i) &\leftarrow Fires(S_{i-1}) \vee Fires(S_{i+1}). \\ &\dots \\ Fires(S_n) &\leftarrow Fires(S_{n-1}). \end{aligned}$$

In this case, one can still defend the PCM solution as an accurate picture of reality: even though the problem description does not mention it, the propagation of shots down the firing line would in fact not be instantaneous, since each soldier takes some time to fire. However, this is not always possible. Consider, for instance, a bicycle which has a big and a small gear wheel connected by a chain. If we turn one of these wheels, the other will turn too. Moreover, this effect will be instantaneous; there is no perceptible delay between turning one and seeing the other turn. Again, CP-logic handles this fine:

$$\begin{aligned} Turn(Big) &\leftarrow Turn(Small). & Turn(Big) &\leftarrow Peddle. \\ Turn(Small) &\leftarrow Turn(Big). \end{aligned}$$

The reader can easily check that this behaves correctly in the face of all conceivable interventions (e.g., manually turning a gear wheel, removing the chain, blocking a gear wheel).

To represent this system as a PCM, we would need to perform the same trick as before:

$$Turn_2(Big) := Turn_1(Big) \vee Turn_1(Small). \quad (8)$$

$$Turn_2(Small) := Turn_1(Big) \vee Turn_1(Small). \quad (9)$$

$$Turn_1(Big) := Peddle \quad (10)$$

However, the newly introduced RVs are now truly artificial: one can no longer explain the difference between $Turn_1(Big)$ and $Turn_2(Big)$ in real-world terms, because, unlike in the firing squad, it cannot be the case that these RVs refer to the condition of the same gear wheel at different points in time. In this case, $Turn_2$ will tell us which wheels will turn, while $Turn_1$ has no real-world meaning. The need to invent artificial RVs in order to model the perfectly intuitive causal relation between these two gear wheels makes this an inadequate representation.

6 More interventions

As Pearl’s book explains, one of the reasons for wanting formal definitions of interventions and their effects is that human experts tend to misjudge such things. Thanks to Pearl, however, such judgments are no longer needed: all you need to do is (1) come up with a causal model of the domain, and (2) figure out how to formulate the intervention you want to consider in terms of the vocabulary of this causal model. All else, i.e., actually figuring out the effect of the intervention, can then be left up to the formal definitions.

One obvious limitation of this methodology is that it is of course not necessarily possible to formulate the intervention you want to consider in terms of the vocabulary of the causal model. In such a case, you cannot blindly let the definitions do the work, but you must still get actively involved and make some changes to the original model. While this cannot be avoided, we may hope to make this need for manual intervention as small as possible. That is, we would like our causal models and formal tools to allow as many reasonable interventions as possible to be applied “unthinkingly”, without the need to tinker with the original causal model.

Some examples of reasonable interventions and associated counterfactuals:

- If we were to remove the chain from our bicycle, peddling would still cause the big gear wheel to turn, but the small wheel would no longer turn with it. Therefore, given that you originally were peddling, the big wheel would still have been turning, even if you had removed the chain.
- Suppose we could send soldier A to some additional training, which decreases his nervousness from q to $q/2$. If we had done this, the probability of the prisoner dying would have dropped from $p + (1 - p)q$ to $p + (1 - p)\frac{q}{2}$.
- In the example of the n person firing squad, suppose that A fired out of nervousness, which caused all other rifleman to fire as well. Would this still have happened if we had somehow managed to make our soldiers a little more stress resistant, such that they would only have fired themselves if *both* their neighbours fired, instead of at least one?

In CP-logic, each of these three examples corresponds to a straightforward intervention with the original theory, namely:

$$\begin{aligned} & \left(\{ \text{Turn}(\text{Big}) \leftarrow \text{Turn}(\text{Small}), \text{Turn}(\text{Small}) \leftarrow \text{Turn}(\text{Big}) \}, \quad \{\} \right) \\ & \left(\{ \text{Fires}(A) : q \leftarrow \}, \quad \{ \text{Fires}(A) : \frac{q}{2} \leftarrow \} \right) \\ & \left(\{ \text{Fires}(S_i) \leftarrow \text{Fires}(S_{i-1}) \vee \text{Fires}(S_{s+1}) \mid 1 < i < n \}, \right. \\ & \quad \left. \{ \text{Fires}(S_i) \leftarrow \text{Fires}(S_{i-1}) \wedge \text{Fires}(S_{s+1}) \mid 1 < i < n \} \right) \end{aligned}$$

Pearl, however, considers only interventions that replace the defining equations for some RVs \mathbf{X} by truth assignments $\mathbf{X} = \mathbf{x}$. If we start from the PCMs for these example as we gave them in the previous section, then none of the three interventions listed above actually corresponds to such an intervention $\mathbf{X} = \mathbf{x}$:

- To remove the effect of the small gear wheel on the big one, Pearl would have to preempt the equation that defines whether the big wheel turns (equation (8)). However, this also removes the effect of peddling.
- Because Pearl’s interventions fix RVs to a specific value, they cannot contain a probability distribution.
- For the same reason, they can also not introduce a new relation between existing RVs.

In all of these cases, the fix is to somehow already include the intervention that we wish to perform in the model. For instance, for the second case:

$$\begin{aligned} \text{Fires}(A) := & \text{Captain} \vee \text{NervousWithoutTraining} \wedge \neg \text{Training} \\ & \vee \text{NervousWithTraining} \wedge \text{Training}. \end{aligned}$$

In itself, this is not hard, but for the reasons outlined at the beginning of this section, the CP-logic way of handling such interventions without changing the original causal model is preferable.

7 Related work

Shortcomings of PCMs have already been recognized in the literature. For instance, Hopkins and Pearl [6] join us in observing that: “In structural causal models, everything is represented as a random variable. Thus, one cannot distinguish between an enduring condition (e.g. the man is dead) versus a transitional event (e.g. the man dies).” They then attempt to fix this and other problems, by means of Situation Calculus. To us, this seems like overkill. SitCalc is an expressive action language, which contains many features that go beyond what is traditionally expressed in a causal model (e.g., preconditions for actions, fluents that might spontaneously change value, and frame axioms for prohibiting fluents from spontaneously changing values). For typical causal reasoning problems, these features are not needed: for instance, one is always free to consider

any intervention (= action) whatsoever, and the value of the fluents is fully bound by the causal laws. In any case, Hopkins’ approach requires that all “intervenable” properties be represented as SitCalc actions, which means that (1) it cannot handle causal cycles such as the gear wheels any better than regular PCMs (Section 5), and (2) it is equally limited in the kind of interventions that can be considered without changing the original model (Section 6).

The semantics of CP-logic is closely related to the well-founded model construction of logic programming (LP). There are a number of other LP languages that deal with probability and causality, on which we will now briefly comment. We discuss only issues related to the specific topic of this paper (interventions and counterfactuals); for a more general comparison, we refer to [10].

P-log [1] is a language that performs probabilistic reasoning with answer sets. It comes equipped with a *do*-operator for performing interventions, which, as shown in [2], can be used to perform counterfactual reasoning in P-log. Essentially, this *do*-operator is the same as Pearl’s. Therefore, we could repeat here the comments made earlier. P-log translates its causal models, together with interventions and observations, to an ASP program, which is then combined with an ASP knowledge base. Instead of relying on the *do*-operator, one can also update such a program by adding additional ASP rules. These “non-monotonic updates”, as they are called, provide a much more flexible system for interventions, but they do require knowledge of how the high-level probabilistic construct are translated into the ASP encoding. A second similarity between P-log and Pearl is P-log’s *coherence* criterion, which is similar to Pearl’s condition that an assignment of values to the exogenous variables should uniquely determine the values of the endogenous ones. Moreover, like Pearl, Baral et al. suggest ensuring this criterion by means of an acyclicity condition. The arguments we gave earlier regarding the advantages of CP-logic when it comes to representing cyclic causal relations also carry over to P-log.

As shown in [10], Poole’s Independent Choice Logic (ICL) [8] is a sublanguage of CP-logic, to which the whole logic can be mapped in a polynomial and modular way. [3] examines interventions (as well as related notions such as actual causes and explanations) in this logic. However, they arrive at their definitions by means of a transformation into Pearl’s causal models. This is the opposite of our approach: they take the fine-grained, event-based representation (restricted to the acyclic case), compile it into the coarser RV-based representation, and then do the interventions there. As we have argued above, there are significant advantages to defining interventions directly on CP-laws.

8 Conclusions

This paper has presented an analysis of interventions and counterfactuals, that reformulates Pearl’s intuitions in the Shaferian framework of CP-logic. Our treatment has some attractive features: we can elegantly handle cyclic causality and can deal with several kinds of interventions that, for various reasons, can only be handled by Pearl at the cost of tinkering with the original model.

Besides these practical advantages, our work also makes philosophical contributions. First, it ties together different approaches to causality from the literature: we investigate Pearl’s interventions and counterfactuals in Shafer’s semantic context, using the LP-based language of CP-logic to syntactically describe classes of probability trees.

Second, we also add a touch of clarity to the picture painted by Pearl. His book considers causal models in two different languages: Bayesian networks and PCMs. While they are formally very similar—every Bayesian network is easily transformed into a PCM—they embody views on the nature of causality that are ontologically quite different: Bayesian networks represent causal relations as inherently probabilistic, while PCMs express the *Laplacian* view that causal relations are completely deterministic and uncertainty stems solely from a lack of knowledge about their “inputs”. Pearl’s book adopts Bayesian networks throughout the chapters that first introduce the idea of interventions. When eventually the topic of counterfactuals arises, however, a switch is made to the Laplacian view of PCMs. It is peculiar that, on the one hand, interventions should be easiest to explain under the assumption that causal relations are inherently probabilistic, while on the other hand, their use for counterfactual reasoning requires the assumption that causal relations are deterministic. Our paper shows that CP-logic’s event-based view on causality reconciles these two views: *whether* a Humean event happens is deterministic (in any given state of the world), but its *outcome* can be probabilistic. In this way, CP-logic can match Bayesian networks as a natural representation for probabilistic causal relations, while also, as this paper has shown, surpassing PCMs as a counterfactual reasoning tool.

References

1. C. Baral, M. Gelfond, and N. Rushton. Probabilistic reasoning with answer sets. *Theory and Practice of Logic Programming*, 9(1), 2009.
2. C. Baral and M. Hunsaker. Using the probabilistic logic programming language P-log for causal and counterfactual reasoning and non-naive conditioning. In *Proceedings of IJCAI*, 2007.
3. A. Finzi and T. Lukasiewicz. Structure-based causes and explanations in the independent choice logic. In *Uncertainty in Artificial Intelligence, UAI*, 2003.
4. J. Halpern and J. Pearl. Causes and explanations: A structural model approach – part I: Causes. In *Uncertainty in Artificial Intelligence, UAI*, 2001.
5. J. Halpern and J. Pearl. Causes and explanations: A structural-model approach. part I: Causes. *The British Journal for the Philosophy of Science*, 56(4), 2005.
6. Mark Hopkins and Judea Pearl. Causality and counterfactuals in the situation calculus. *J. Log. Comput.*, 17(5):939–953, 2007.
7. J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge Press, 2000.
8. D. Poole. The Independent Choice Logic for modelling multiple agents under uncertainty. *Artificial Intelligence*, 94(1-2):7–56, 1997.
9. G. Shafer. *The art of causal conjecture*. MIT Press, 1996.
10. J. Vennekens, M. Denecker, and M. Bruynooghe. CP-logic: A language of probabilistic causal laws and its relation to logic programming. *Theory and Practice of Logic Programming*, 9(3):245–308, 2009.